“BIBLIOTECAS Y NUEVAS LECTURAS EN EL ESPACIO DIGITAL”

Buenos Aires 14 al 17 de abril de 2007

Miércoles 18 de abril de 2007 - Sala Roberto Arlt

Seminario

Primer proyecto de biblioteca digital: ¿por donde comenzar?

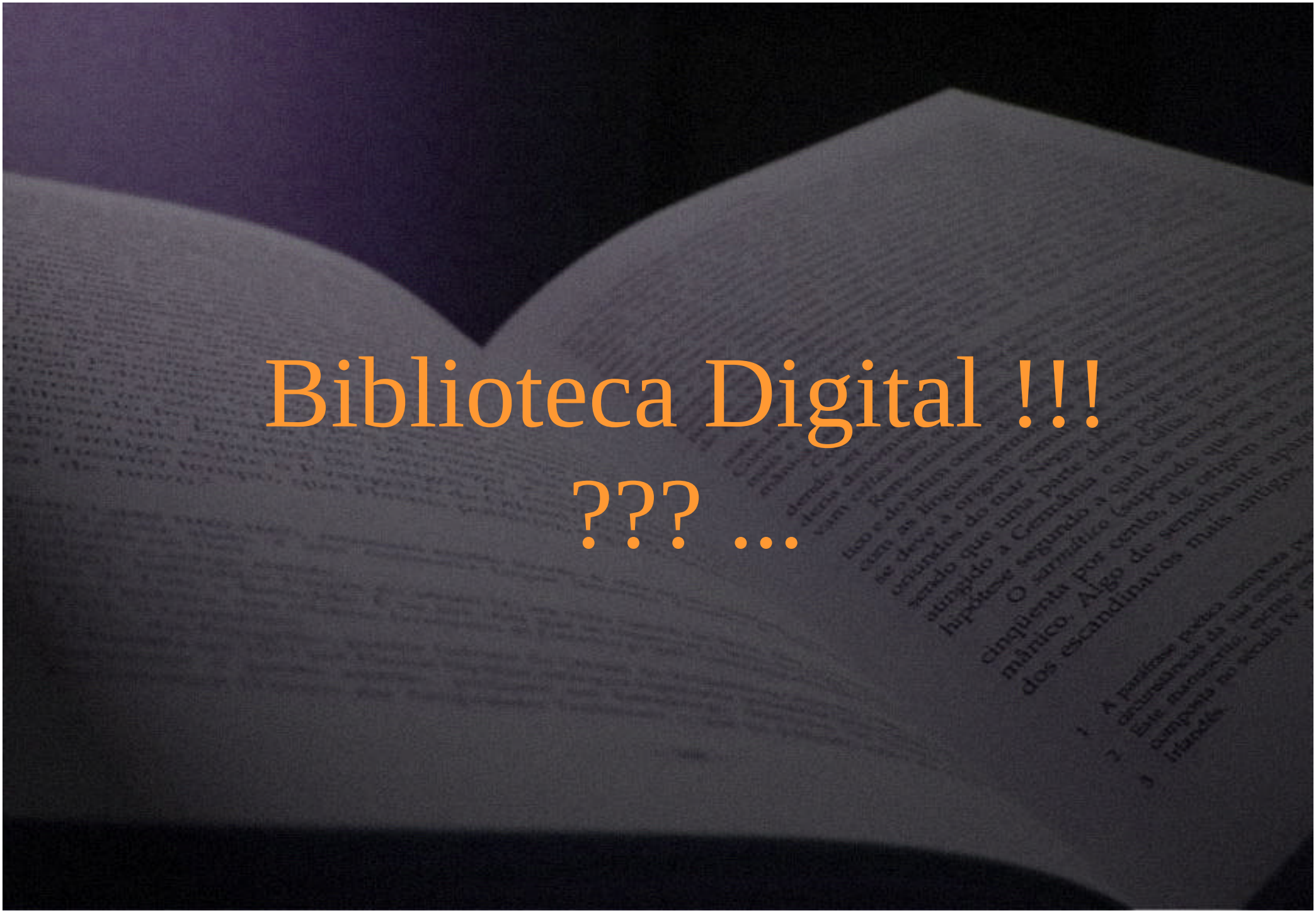
Diego Spano

ABGRA

Primer proyecto de Biblioteca Digital: ¿por dónde empezar?

2.º Congreso Iberoamericano de Bibliotecología
"Bibliotecas y nuevas lecturas en el espacio digital"

14 al 17 de abril de 2007
La Rural - Predio Ferial de Buenos Aires


Biblioteca Digital !!!
??? ...

•Qué es una “biblioteca digital”?

Es una colección de objetos digitales (texto, video, audio) junto a métodos de acceso y recuperación locales o remotos.

Antes de empezar…

- ¿Cuál es el objetivo de mi colección?
- ¿A qué grupo de usuarios se dirige?
- ¿Cuán grande es dicho grupo? ¿Tiene dimensión local, regional o mundial?
- ¿Cuántos documentos tengo pensado incluir en la colección?
- ¿Cuántas páginas?
- ¿Cuánta información gráfica contienen?
- ¿Cabe subdividir la documentación en partes que sean de interés para un público
- ¿Existen ya en forma electrónica los documentos?
- De ser así, ¿en qué formato?

Antes de empezar…

- ¿A qué derechos de autor están sujetos los documentos?
- ¿Quién es el titular de los derechos de autor?
- ¿Hay otras organizaciones que se dirijan al mismo público?
- ¿Tengo deseos de colaborar con otros grupos?
- ¿De qué presupuesto dispongo para el conjunto del proyecto?
- ¿De qué recursos humanos (en personas-mes) dispongo para las labores de coordinación,
- ¿De cuántas computadoras dispongo para el proyecto?
- ¿Cuántos CD-ROM tengo pensado distribuir?
- ¿Voy a distribuirlos gratuitamente o a venderlos?

•Qué son los objetos digitales?

Elementos constituídos por:

CONTENIDO

+

METADATOS

•Para qué sirven las BD?

- Administración de conocimientos/contenidos
  - Administrar y acceder a repositorios de información internos
- Comunicación, educación, investigación
  - E-journals, e-prints, e-books, e-learning
- Acceso a colecciones culturales
- Archivo y preservación
- …

Software BD: requerimientos claves

- Tipos de documentos (libros, periódicos, páginas web, …)
- Formatos (texto, PDF, Word, PS, HTML …)
- Adquisición de contenidos (online y offline)
- Indexación y recuperación
  - Indexación estructurada / full text
  - Extracción automática de metadatos
- Almacenamiento
  - Compresión de datos
  - Almacenamiento eficiente de metadatos
  - Ubicación rápida de metadatos y documentos
- Acceso y entrega
  - Búsquedas estructuradas, no estructuradas y navegación jerárquica
  - Distribución en CD/DVD

Software BD: Más requerimientos

- Escalable para grandes colecciones
- Soporte Multilingual
- Administración de seguridad de acceso
- Monitoreo y reporte de utilización
- Conformidad con estándares
  - XML, Dublin Core, Unicode
- Interoperación
  - OAI, Z39.50, MARC, CDS/ISIS, …

•A qué se apunta…

- “Colecciones” de material digital
- Hasta varios Gb de texto …
- … + imágenes asociadas, películas, mp3, etc, etc
- Totalmente “indexada”
- Servida en WWW, o publicada en CD-ROM
- Multi-plataforma (Unix + Windows + Mac)
- Multi-formato de documentos y metadata
- Multi-lingual: documentos e interfases
- Multimedia
- Metadatos: estándar y no-estándar

•Software BD: Alternativas

- Cuáles son las expectativas?
- Desarrollo de una aplicación web local?
- Solución comercial?
- Adoptar software open source?
  - Greenstone
  - Eprints
  - DSpace
  - (CDS/ISIS, Koha)

GreenStone
Software para la construcción de
bibliotecas digitales

•Qué es Greenstone?

- Greenstone ha sido elaborado como parte del proyecto de Biblioteca Digital de Nueva Zelanda por la Universidad de Waikato y actualmente es desarrollado y distribuido en colaboración con la UNESCO y la ONG Human Info.

- Es un software de código abierto disponible en ***http://www.greenstone.org*** bajo los términos y condiciones de la Licencia Pública General de GNU.

•Qué es Greenstone?

- Paquete de software para crear, mantener y distribuir colecciones digitales
- Código abierto (Open-source)
- Desarrollado por la Universida de Waikato
- Socios para su distribución:
  - UNESCO
  - Human Info NGO, Bélgica
  - Todos!

•Qué es Código Abierto?

- “La idea básica detrás de ‘código abierto’ es muy simple:

Cuando los programadores pueden leer, redistribuir y modificar el código fuente de un programa, entonces el software evoluciona. La gente lo mejora, lo adapta, arregla fallas. Y todo esto pasa a una velocidad que, aquél que está acostumbrado a los tiempos convencionales de desarrollo de software queda asombrado.“

(www.opensource.org)

- Cualquiera puede redistribuir el software, aun bajo pago.

- El código fuente de estar siempre disponible.

•Greenstone, bibliotecas y acceso abierto

"El propósito del software es alentar a los usuarios, particularmente en universidades, bibliotecas y otras instituciones de servicio público a construir sus propias bibliotecas digitales.
Estas están cambiando radicalmente la manera en la que la información se está adquiriendo y diseminando entre las comunidades e instituciones socias de la UNESCO, en los campos de Educación, Ciencia y Cultura alrededor del mundo, y particularmente en países en desarrollo.

Compartir información y hacerla de dominio público, ese es el objetivo."

www.greenstone.org

•Greenstone

Acceso

- Accesible via Web browser
- El Servidor corre en Windows y Unix
- Colecciones publicadas en CD-ROM

Búsqueda/ navegación

- Full-text y por campos
- Opciones flexibles de navegación
- Basado en Metadata (Dublin Core)
- Independiente en cada colección
- Búsqueda jerárquica de frases

Extensible

- Plugins — documento nuevo, formatos de metadatos
- Clasificadores — creación libre

Multilingual

- Documentos e interfases
- Chino, Arabe, Maori, Ruso, etc
- Multimedia: video, audio

•Dónde encontrarlo

- Disponible en: www.greenstone.org
  - Software
  - Colecciones de demo
  - FAQ
  - Ejemplos
  - Colecciones documentadas
- Documentación:
  - Guía de instalación, Guía del usuario, Guía del desarrollador, Del papel a la colección
- Lista de Mailing:
  - Lista de usuarios de Greenstone
  - Lista de desarrolladores de Greenstone

•Plataformas

- Sistemas operativos:
  - Windows (+95)
  - Linux (cualquier versión)
  - Unix
  - Mac OS X (algunos problemas con GLI)
- Restricciones:
  - No corre sobre Windows 3.1/3.11
  - Para la Interfase de Bibliotecario (GLI) se necesita Java— el cual no está más soportado en Windows 95
- Espacio en disco
  - 50 MB para el programa
  - 250 MB para la colección de demo (opcional)
  - 25 MB para la función “exportar a CD”

•Biblioteca Local

- El paquete de instalación de Windows contiene todos los componentes necesarios para instalar Greenstone con la colección de demo.

- Directorio de instalación por defecto:
  C:\archivos de programa\greenstone

- Levantar la aplicación:
  Inicio → Greenstone digital library

- Usar la Interfase de Bibliotecario(GLI):
  Inicio → Greenstone digital library

•Biblioteca Local vs Biblioteca Web

- Biblioteca local: standalone
  - Sirve colecciones en una PC …
  - Y a otros en una misma red
  - Incluye un servidor Web integrado

- Biblioteca Web: usa un server web externo
  - Apache, Microsoft PWS/IIS

•Pasos en el proceso de construcción de una colección

Nº 1 Definición de la colección.

Nº 2 Configuración de la colección.

Nº 3 Selección e importación de los documentos.

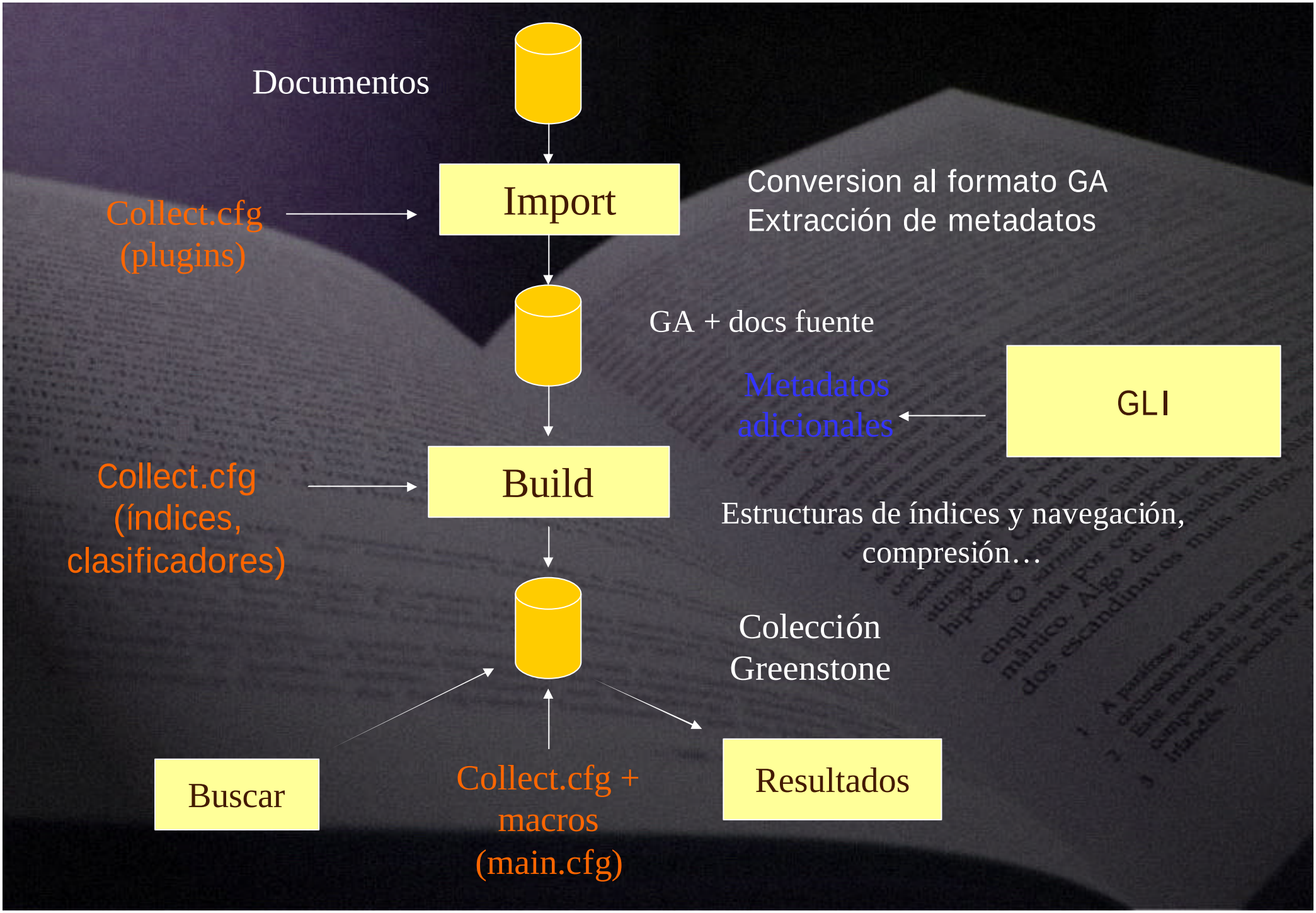
Nº 4 Descripción de los documentos.

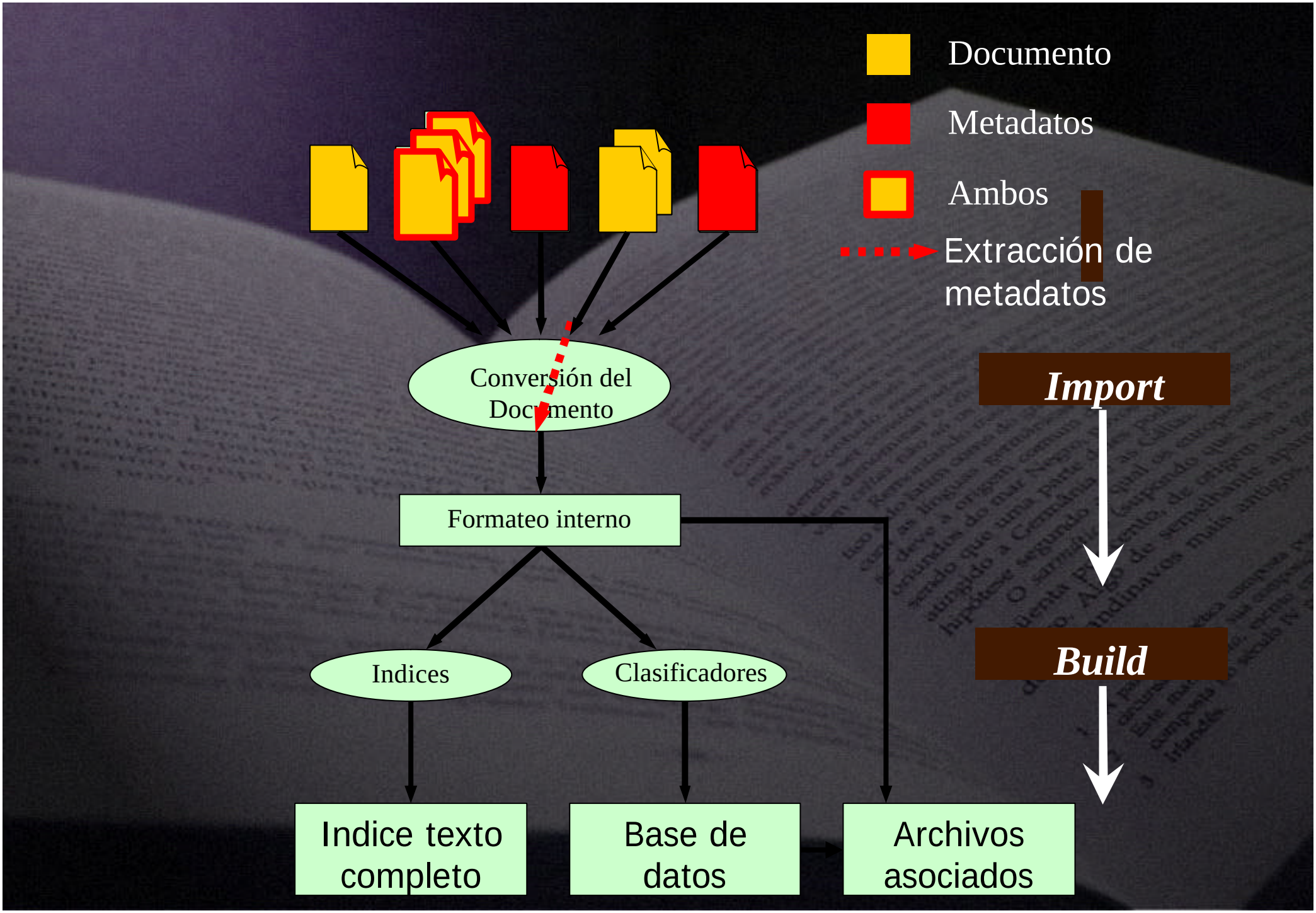
Nº 5 Construcción de la colección.

Nº 6 Puesta en servicio / Recuperación de la información.

•Proceso de construcción de una colección

- Input: un conjunto de documentos
- Greenstone “importa” esos documentos y los convierte al formato GA (XML/HTML)
- Greenstone “construye” los índices y las estructuras de navegación usando los archivos GA
- El archivo de configuración de la colección determina la conversión del contenido, la extracción y construcción de índices y los clasificadores
- La presentación de los resultados de las búsquedas y/o clasificadores y de la interfase se determina con las cadenas de formateo y las macros


Documentos
Import
Collect.cfg
(plugins)
Conversion al formato GA
Extracción de metadatos
GA + docs fuente
Metadatos
adicionales
GLI
Collect.cfg
(índices,
clasificadores)
Build
Estructuras de índices y navegación,
compresión…
Colección
Greenstone
Buscar
Collect.cfg +
macros
(main.cfg)
Resultados


Documento
Metadatos
Ambos
Extracción de metadatos
Conversión del Documento
Formateo interno
Indices
Clasificadores
Indice texto completo
Base de datos
Archivos asociados
Import
Build

•Modelo jerárquico de documentos

Texto Expandir

Juventud e Identidad - Tomo II

- Indice
- Prólogo
- Capitulo I: El congreso. Acto de apertura (25-IX-97)
- Capitulo II: Labor de las Filiales
- Capitulo III: El derecho a la identidad
- Capitulo IV: Identidad e Investigación Genética
- Capitulo V: Identidad y antropologia forense
- Capitulo VII: La lucha permanente por la restitución situación actual y perspectivas

Haga click aquí para descargar el libro completo

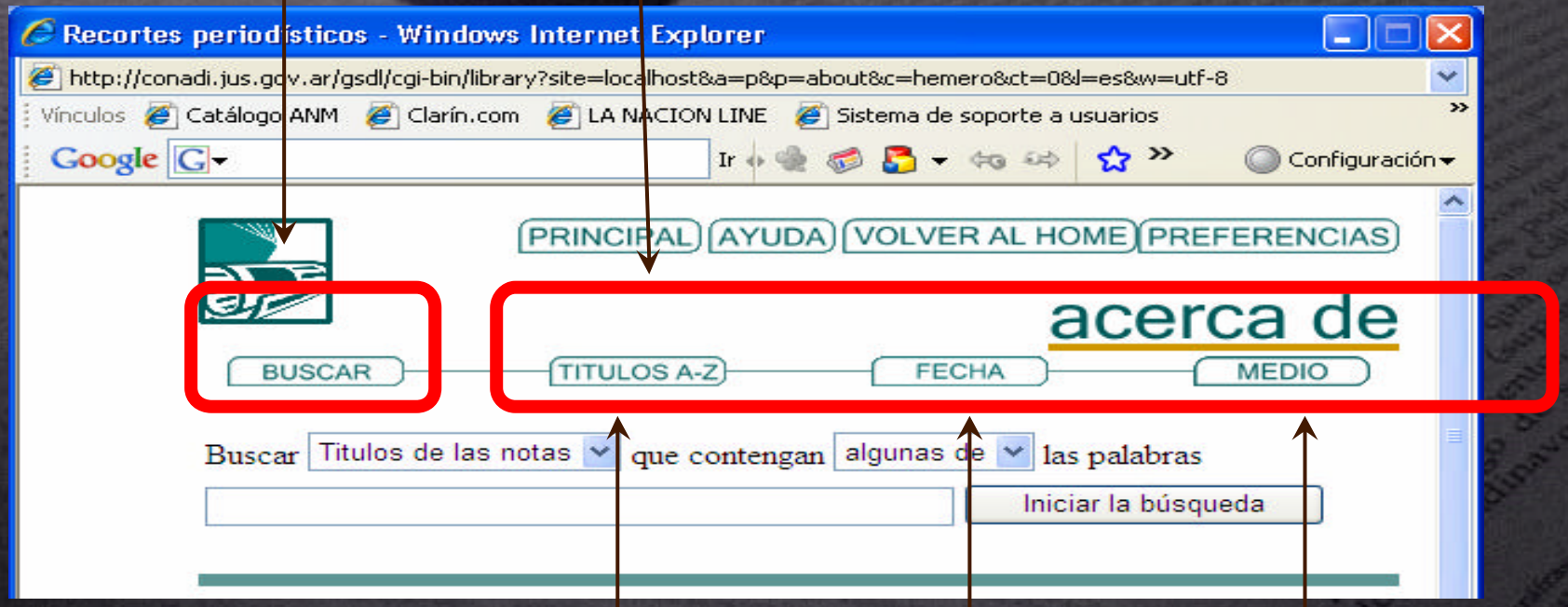
Juventud e Identidad

❖Metadatos especificados a cualquier nivel

•Búsqueda y navegación

❖Búsqueda

❖Navegación por metadatos

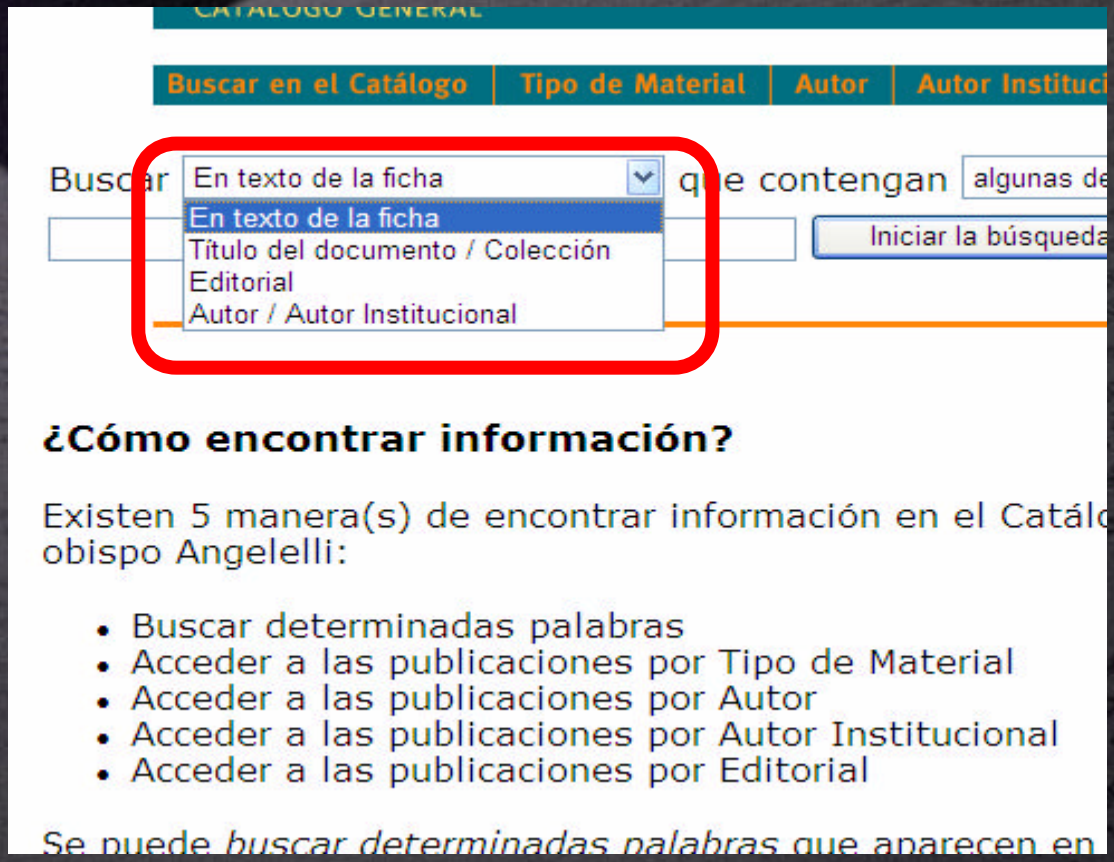


Titulo Fecha Medio

Dublin Core / ad hoc

•Múltiples índices de búsqueda

Independientes entre una colección y otra


Buscar en el Catálogo
Tipo de Material
Autor
Autor Institucional
Buscar
En texto de la ficha
que contengan
algunas de
En texto de la ficha
Título del documento / Colección
Editorial
Autor / Autor Institucional
Iniciar la búsqueda
¿Cómo encontrar información?
Existen 5 manera(s) de encontrar información en el Catálo
obispo Angelelli:
Buscar determinadas palabras
Acceder a las publicaciones por Tipo de Material
Acceder a las publicaciones por Autor
Acceder a las publicaciones por Autor Institucional
Acceder a las publicaciones por Editorial
Se puede buscar determinadas palabras que aparecen en

•Clasificador AZList

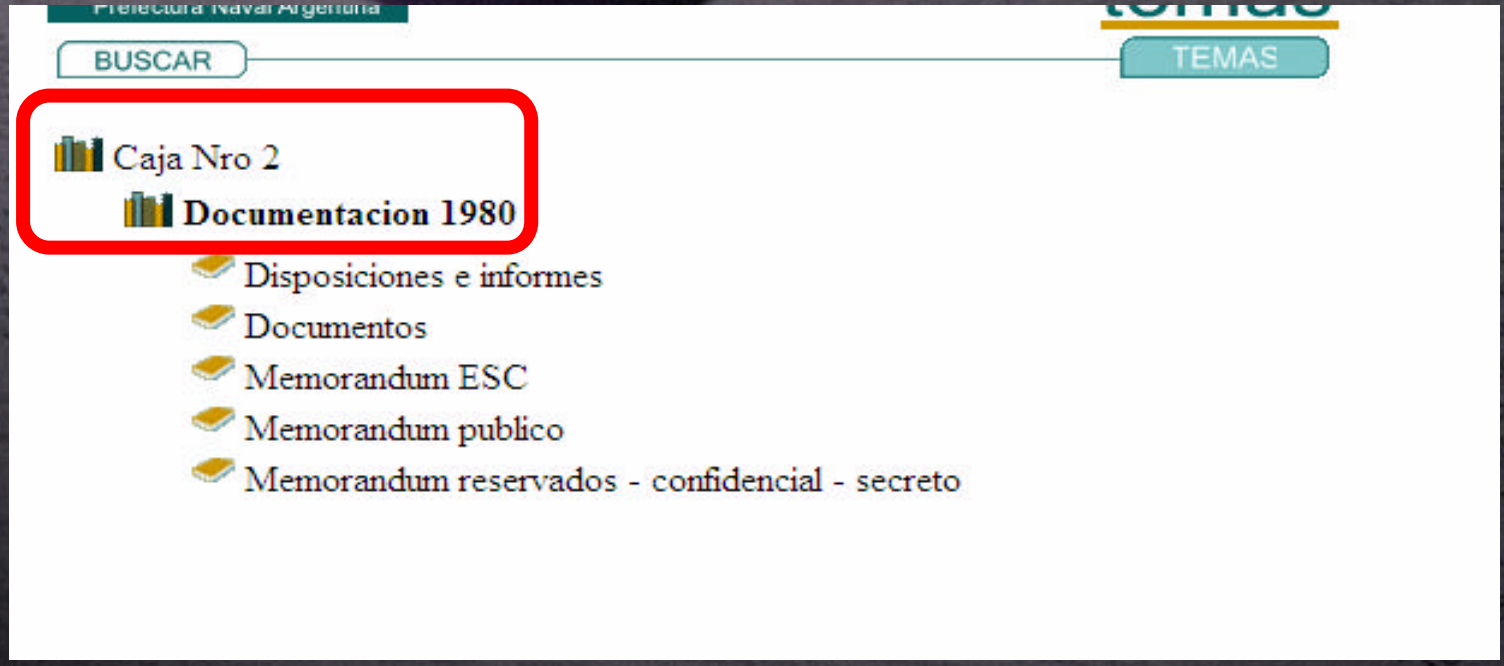
títulos A-Z
BUSCAR
TITULOS A-Z
A B C D E F G H I J K L M N Ñ O P Q R S T U V W X Y Z ...
Carla Graciela Rutila ARTES
José Sabino ABDALA
Laura Fernanda ACOSTA
Maria F. ALVAREZ
Pablo Germán ATANASIEU LASCHAN
Paula Noemí ALBORNOZ MUSASCHIO
Tamara Ana Maria ARZE

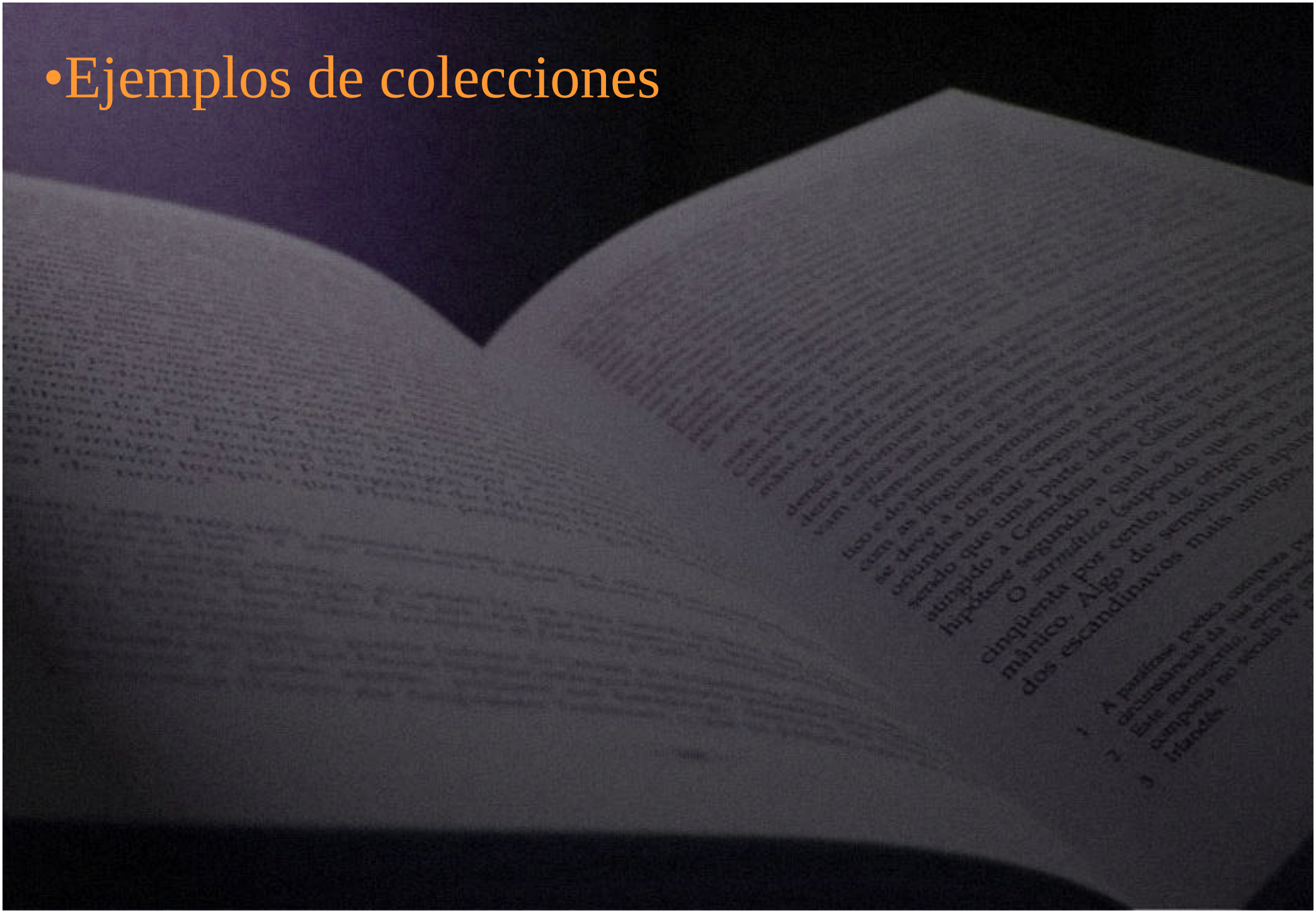
•Clasificador DateList

fechas
BUSCAR
TITULOS A-Z
FECHA
MEDIO
2003 2004
2004 mayo
Catamarca - Las Abuelas de Plaza de Mayo llega a Catamarca - (27 mayo 2004 / Diario El Ancasti)
junio
La Rioja - El trabajo de las Abuelas y de la Conadi - (2 junio 2004 / Diario El Independiente)
La Rioja - Se conformó la red por la identidad - (2 junio 2004 / Diario Nueva Rioja)
La Rioja - Forman una red para recuperar la identidad de hijos de desaparecidos - (2 junio 2004 / Diario El Independiente)
Catamarca - Abuelas en Catamarca - (3 junio 2004 / Diario El Ancasti)
Catamarca - Se conformó la red por el derecho a la Identidad - (4 junio 2004 / Diario La Union)
julio
Catamarca - Llegan miembros de Abuelas y de la Conadi - (2 julio 2004 / Diario La Union)
La Rioja - Atenderán a todas las personas que duden de su identidad - (8 julio 2004 / Diario Nueva Rioja)

•Clasificador List

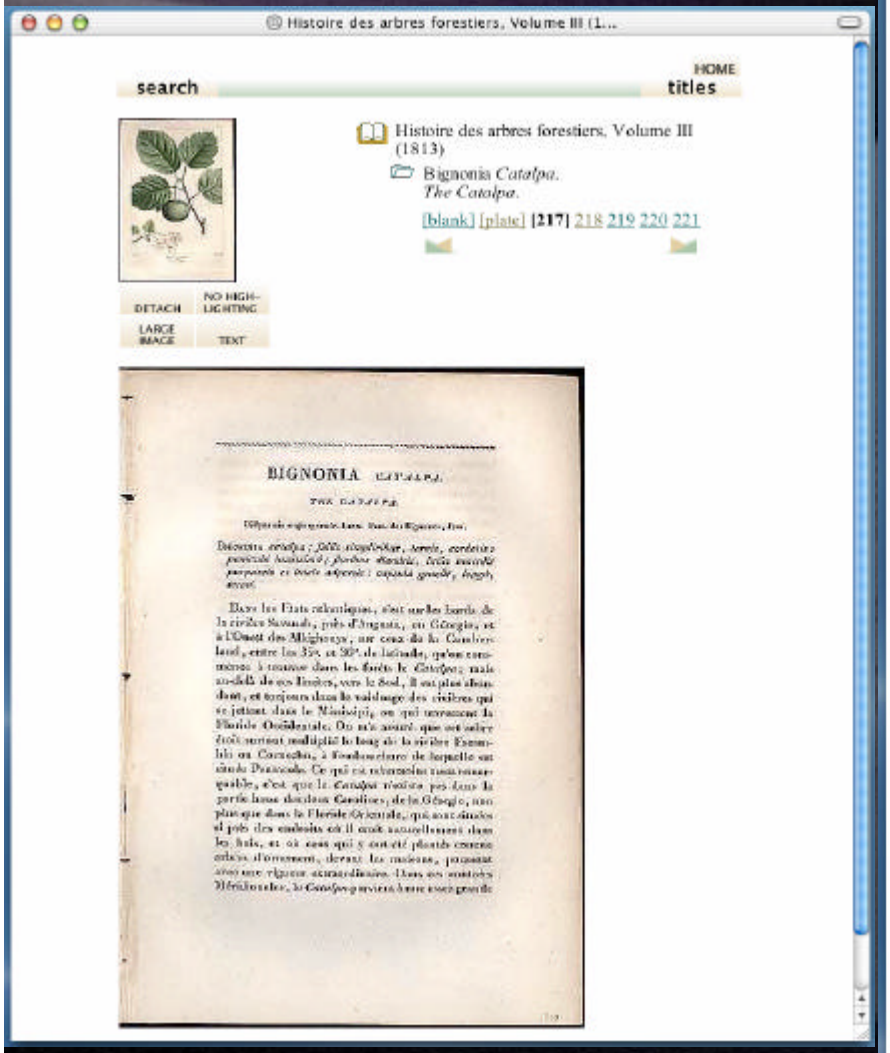
•Clasificador Hierarchy

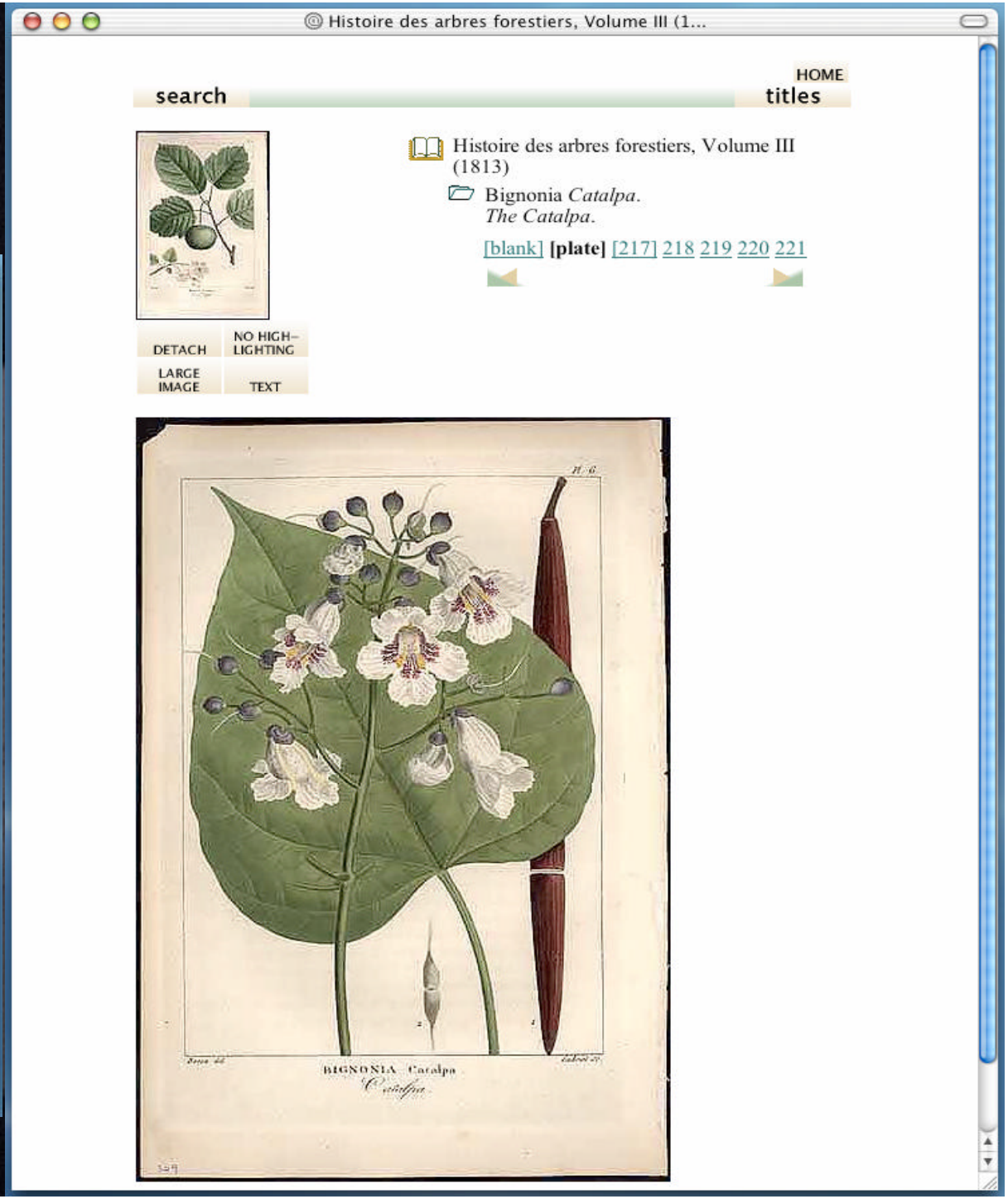


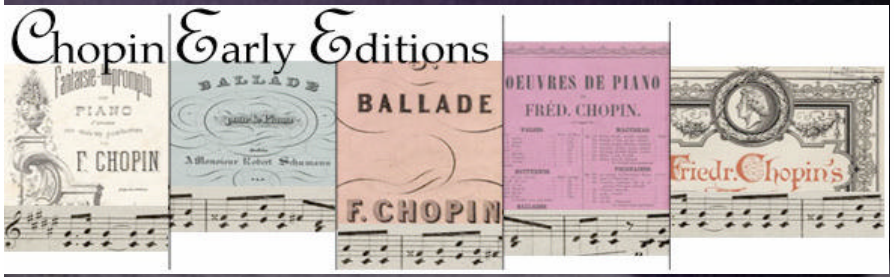


- Ejemplos de colecciones

- **Jardín botánico de New York**





Biblioteca de la Universidad de Chicago

Allegro de concert : pour le piano : op. 46 /

The University of Chic
LIBRA

Chopin Early Editions

HELP PREFER

search titles uniform titles genres dedica

Allegro de concert : pour le piano : op. 46 /

view score | bibliographic description

Title: Allegro de concert : pour le piano : op. 46 /
Composer: Chopin, FrÃ©dÃ©ric, 1810-1849.
Dedicatee: MÃ¼ller-Streicher, Friederike
Publication Info.: Leipzig :Chez Breitkopf & HÃ¤rtel :Paris :Chez M. Schlesinger,[1841]
Description: 23 p. of music ; 34 cm.
Notes: composÃ© et dÃ©diÃ© Ã Mademoiselle F. MÃ¼ller par FrÃ©d. Chopin.
Engraved. Price: 1 Thlr. 6 Ngr.
Also available on the Internet.
Subject: Piano music.
Plate No.: 6651
URI: http://pi.lib.uchicago.edu/1001/dig/chopin/017
Call No.: M25.C54 A41

Allegro de concert : pour le piano : op. 46 /...

Chopin Early Editions

HELP PREFERENCES

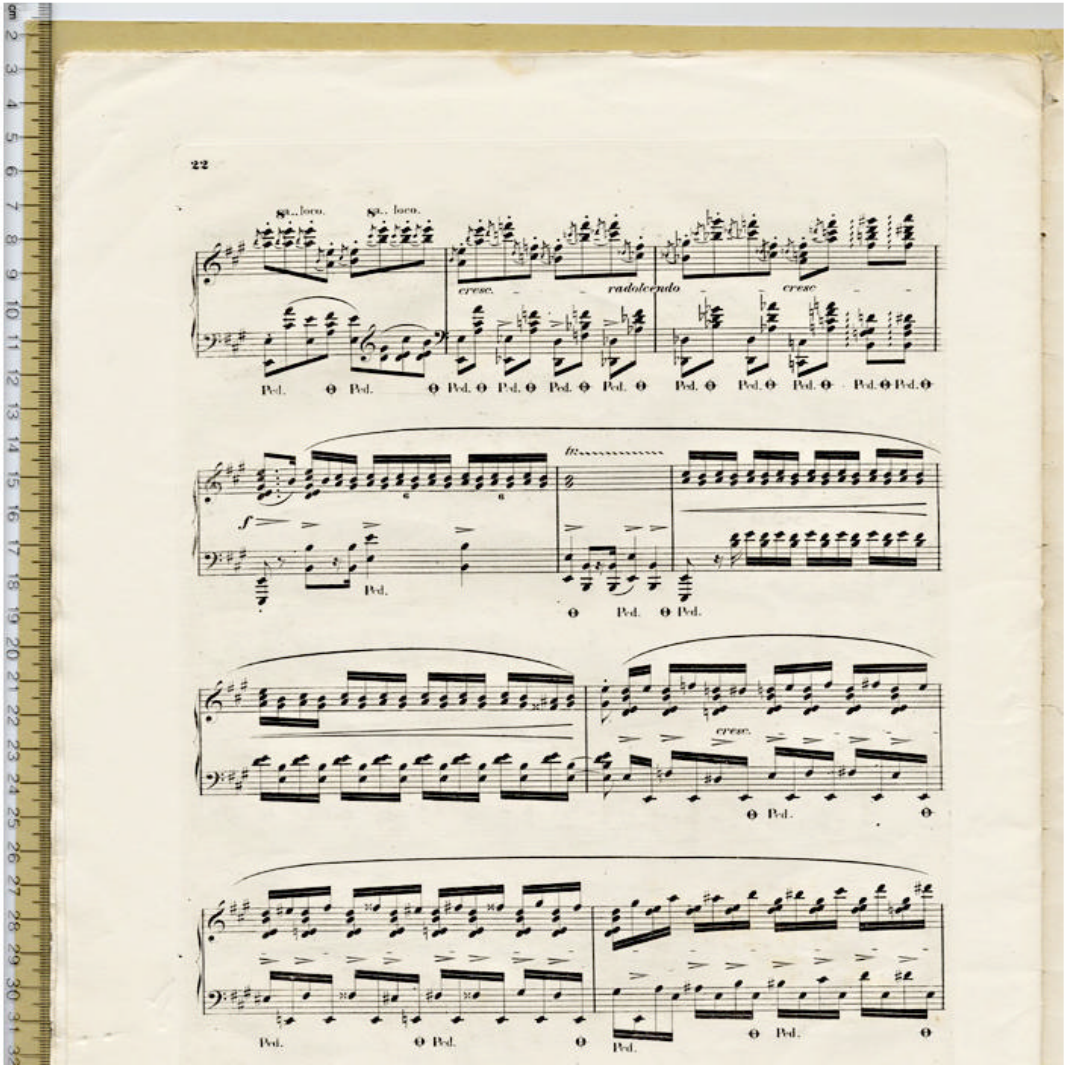
search titles uniform titles genres dedicatees

Allegro de concert : pour le piano : op. 46 /

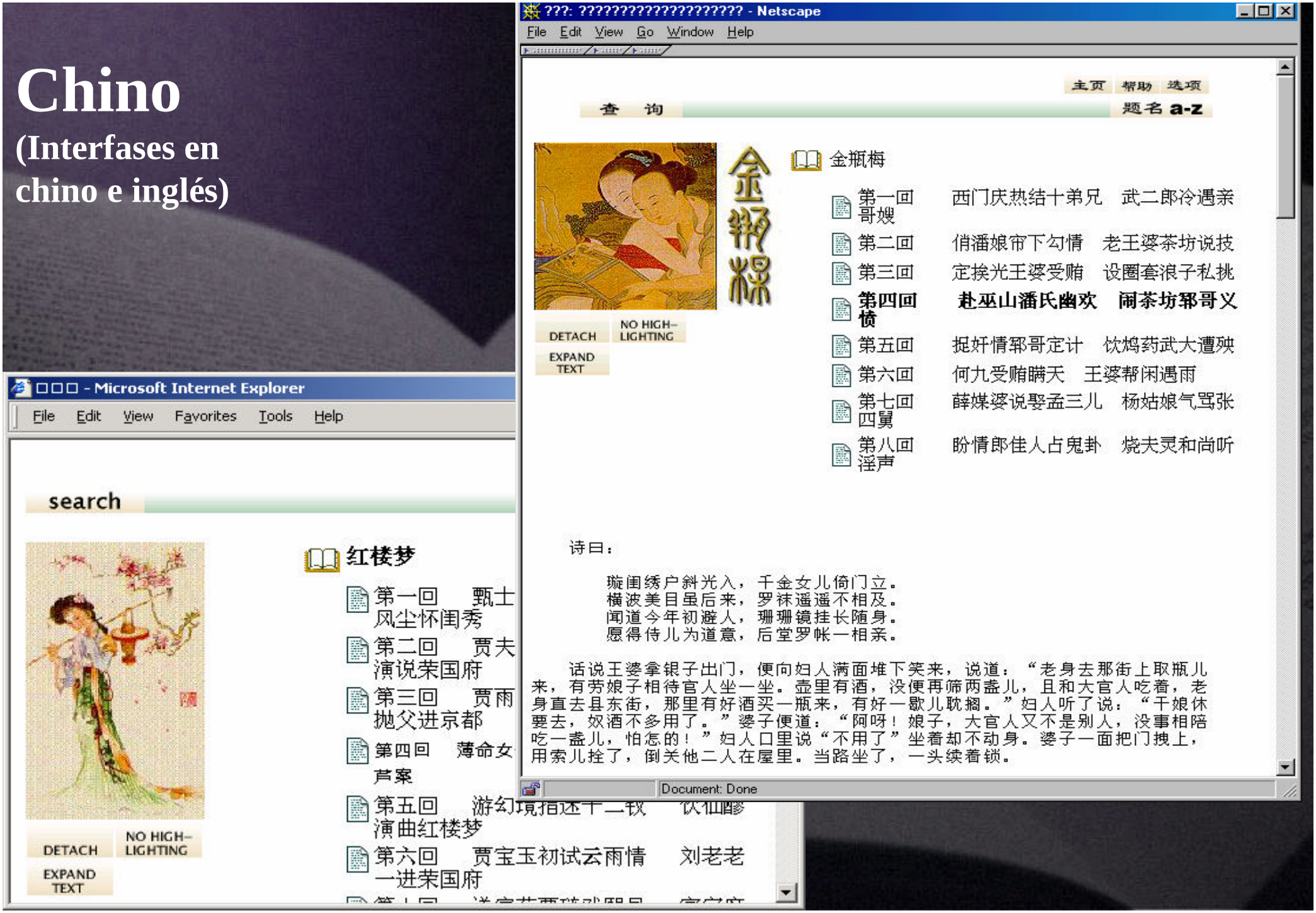
prev next

view score | bibliographic description

Page 22

Page 22 Go

Chino

(Interfases en chino e inglés)



Arabe

العناوين أ-ي - Microsoft Internet Explorer
File Edit View Favorites Tools Help
arabic collection
الموقع مساعدة الخيارات
العناوين أ-ي
ابحث
العناوين أ-ي
جامع الناصر محمد جامع الناصر محمد بن قلاوون شيد
جامع حسن باشا طاهر جامع حسن باشا طاهر انشأه حس
جامع قايتباى جامع قايتباى انشأه الأشرف أبو النصر...
جامع الناصر بالنحاسين جامع ومدرسة السلطان الناصر
مشهد الإمام الجعفرى مشهد الإمام الجعفرى ثم انشأه...
جامع البرديني جامع البرديني انشأه كريم الدين أحمد...
الجامع الحنيق بسوهاج الجامع الحنيق بسوهاج بقع...
الجامع العمرى بقوص الجامع العمرى بقوص بقع بقوص
جامع سلار وسنجر الجاولى جامع سلار وسنجر الجاولى
جامع قانى باى الرماح جامع قانى باى أمير أخور انشأه
السلطان حسن جامع السلطان حسن بدأ فى بنائه السلطان
مشهد السيد يحى الشبيه مشهد السيد يحى الشبيه ثم...
المدرسة الصالحية المدرسة الصالحية انشأها الملك...
جامع الظاهر بيبرس جامع الظاهر بيبرس شيده الظاهر
جامع خاير بك جامع خاير بك بقع بالقاهرة وقام...
عثمان كتخدا جامع عثمان كتخدا يوجد بشارع قصر النيل
جامع الأشرف برسباى جامع الأشرف برسباى بناه السل
جامع الملكة صفية جامع الملكه صفية هذاالجامع انشئ...
جامع برقوق جامع فرج بن برقوق انشأه السلطان الظاهر
جامع الأقمر جامع الأقمر بناه وزير الخليفة الفاطمى...
http://www.nzdl.org/fast-cgi-bin/library?e=q-0arabic--00-prompt-1--1l--1-en-50---20-nav-ÇáÓáØÇä+ - ...
File Edit View Favorites Tools Help
HOME HELP PREFERENCES
search
titles a–z
مدرسة وضريح وبيمارستان السلطان المنصور قلاوون
تم انشاء هذه المجموعة بين عامى 683-684هجرى/ 1284-1285ميلادى وتوجد بشارع المعز لدين الله الفاطمى بمدينة القاهرة
وصف الجامع
اهم جزء فى هذه المجموعة القبة التى تعلو الضريح وتتكون من قاعدة مثمنة مكونة من أربع دعائم مربعة بها أربعة أعمدة رخامية مكسوة من الخارج بالرخام الدقيق المطعم بالصدف وكلا من الدعائم والأعمدة تحمل عقودا مدببة تعلوها رقبة مثمنة ويوجد بكل ضلع منها نافذة ويعلو الرقبة قبة مستديرة والواجهة الخارجية للمجموعة من الحجر الملون الأحمر والأبيض على شكل مربعات والمئذنة تتكون من ثلاثة طوابق الأول والثانى يأخذان شكل مربع والعلوى مستدير وتم تجديد المئذنة على يد الناصر محمد بن قلاوون عام1303 ميلادي وتوجد المدرسة المنصورية امام باب القبة وتم تجديد الإيوان الشرقى بهذه المدرسة بين عامى ميلادى ويتكون من ثلاثة أروقة أكبرها الأوسط والسقف محمول 1916-1919 على أعمدة رخامية تعلوها عقود حليت هى والشبابيك المستديرة فوقها بالزخارف الجصيةاما البيمارستان لم يتبقى

Ruso

Региональные документы

Марийский фольклор

osi

ЗАЩИТА ИНФОРМАЦИИ

Федеральные документы

??????? ?. ?. ?????, ???, ????????: ?????: ?????????? ???? - Netscape

File Edit View Go Communicator Help

начало помощь настройки

поиск названия а-я указатель авторы а-я

Интернет - экономика

весь текст раскрыть оглавление

новое окно снять выделение

Бирюков С. И. Песни, вои, частушки

Песни

Рекуртская песн

Величание

Припев грузчиков

Рекуртская песн

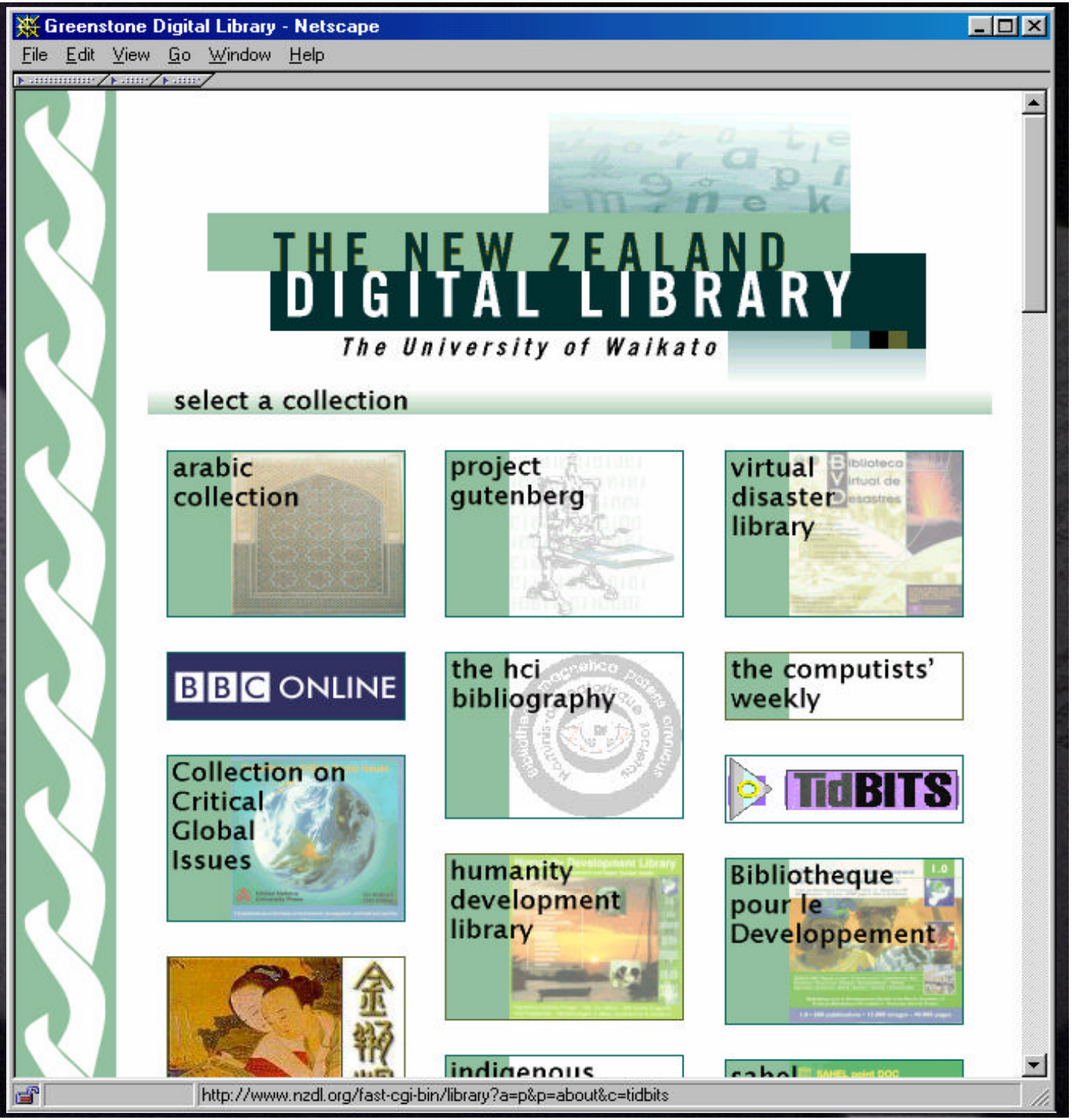
Поется для лица, принятого и уезжающего в армию, накануне его отъезда, группой девушек на вечеринке (беседках).

Как по морю, морю синему,
Посине морю волынчату.
Что плывут, гребут четыре корабля,
А на встречу им червонный корабель.
На корабличке немножко людей,
По моему счету восемь человек,
А девятый на корме сидит,

Document: Done

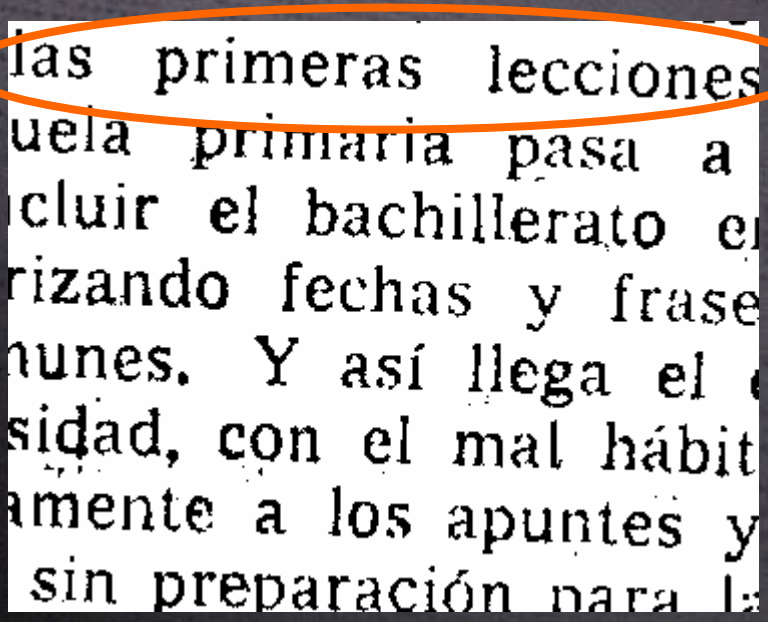
Colecciones en la web

www.nzdl.org



- Scanners y OCR

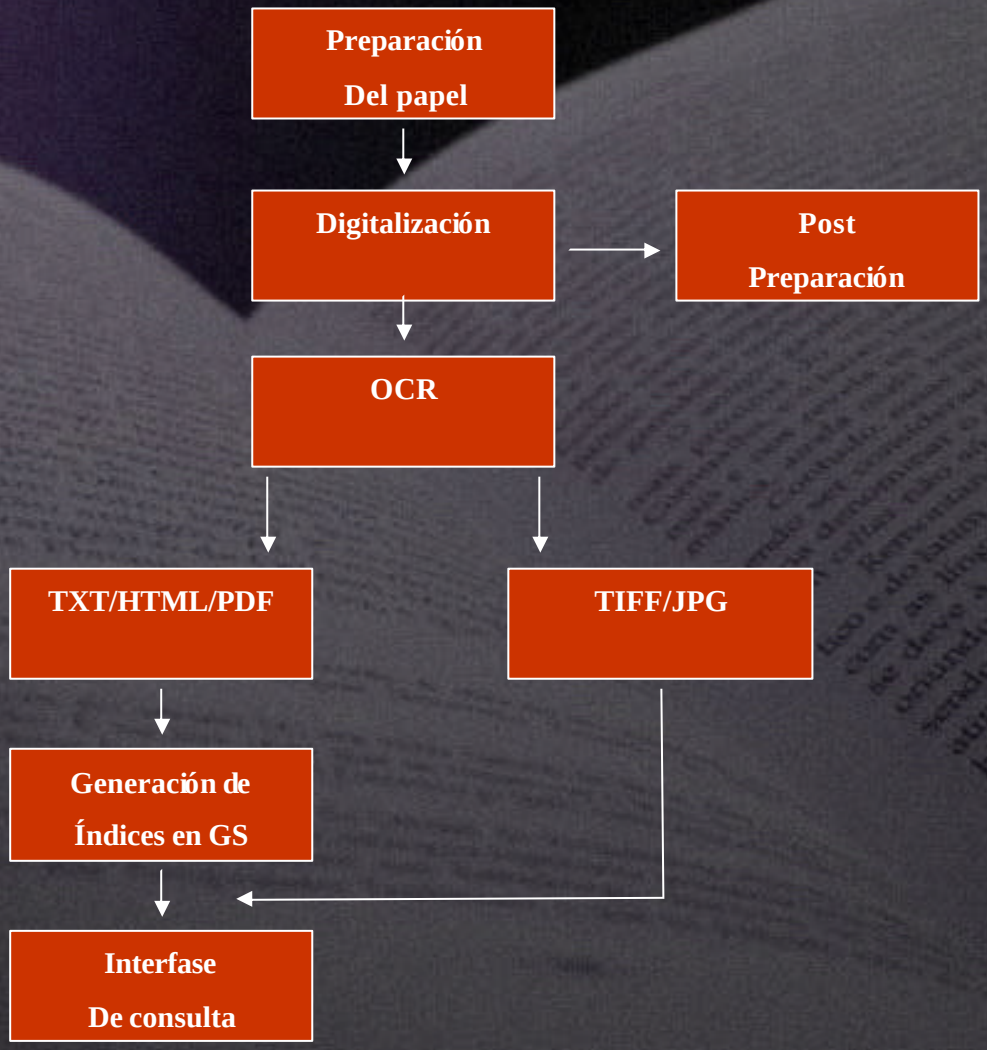
IMAGEN


las primeras lecciones
uela primaria pasa a
cluir el bachillerato e
rizando fechas y frase
nunes. Y así llega el
sidad, con el mal hábit
amente a los apuntes y
sin preparación para la

TEXTO

las primeras lecciones

•Proceso de conversión de docs impresos



•Preparación del papel

Consiste en quitar del papel todo aquél elemento que pueda dañar los espejos del scanner (ganchos, clips) o que pueda interferir en el transporte del papel (roturas, puntas dobladas, pegamento).

Para el caso de libros, puede cortarse el lomo para lograr una correcta digitalización de sus páginas.

•Clasificación de scanners

Por método de alimentación
- Flatbed (cama plana)
- Alimentador automático
- Planetario

- Por método de tracción de papel
  - Rodillos de goma
  - Bomba de vacío
  - Cilindros con bandas

- Por método de captura
  - Simple lado (simplex)
  - Doble lado simultáneo (dúplex)
  - Doble lado no simultáneo

Clasificación de scanners

- Por nivel de producción
  - De escritorio
  - Departamentales
  - Industriales / Alta gama
- Por método de conexión
  - SCSI
  - USB / Firewire
  - Puerto paralelo

•Clasificación de scanners (ejemplos)

Scanner flatbed y adf

Scanner alta producción

Scanner pequeño de tarjetas o pasaportes

Scanner de planos

•Parámetros de escaneo

- Resolución
- Profundidad de color
- Compresión
- Brillo
- Contraste

•Parámetros estándar

- Resolución de escaneo: 200 / 300 dpi
- Formato de archivo
  - Blanco y negro: Tiff
  - Color: Jpeg / Png
- Compresión
  - Blanco y negro: CCITT Grupo III o Grupo IV
  - Color: Jpeg / Png

•Sistemas de OCR

- Read-Iris http://www.readiris.com/
- Omnipage http://www.nuance.com/omnipage/
- Fine-Reader http://www.finereader.com/

•Post-Preparación del papel

Consiste en rearmar el documento y dejarlo en la forma en que estaba originalmente.

En muchos casos se logra que el documento quede incluso restaurado.

Conclusiones

- Una colección básica en Greenstone puede crearse en minutos.
- La posibilidad que tiene Greenstone de ser modificado “a medida” permite cambiar su comportamiento para que se ajuste a las necesidades de cualquier organización.
- Las opciones de Formateo brindan un sinnúmero de alternativas para cambiar la interfase de usuario. Lo mismo ocurre con las ‘macros’ que manejan el comportamiento de la aplicación.
- Greenstone se convertirá en poco tiempo en una herramienta estándar para la creación y administración de bibliotecas digitales.

• Bibliografía

Material de los workshops de Greenstone preparados por Ian H. Witten, David Bainbridge y Michael Dewsnip.

Manual “Del Papel a la Colección”. Greenstone.

Traducido y compilado por Diego Spano.

diegospano@hotmail.com